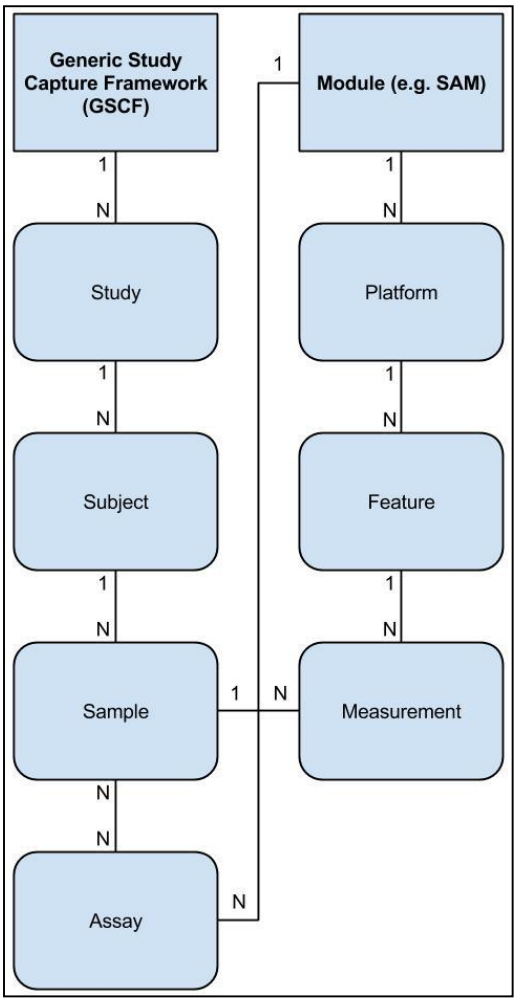
Phenotype Database

Quick start user manual for entering study (meta)data in the Generic Study Capture Framework & Assay Modules

Mariska Bierkens, Jolanda Strubel

1 Introduction

This short manual explains the steps required to create a study in Phenotype Database. It includes the definition of the study, creating of subjects, creating samples and linking data of measurements to these samples using so-called assays. For a full description of Phenotype Database go to the extended user guide available on the website.



Please note before starting with uploading data in Phenotype Database:

- Create templates for each part of the study. Fields within the templates represent data that needs to be entered. Creating templates beforehand will enable the user to directly match their data columns to the particular template fields (see Chapter 3 of the user manual for description on templates).
- Data files that need to be uploaded should adhere to a specific structure to allow matching of column headers to the desired template field (see Paragraph 2.3 of the user manual for some examples).

In this manual a walkthrough will be given, in which a study is set up (Section 2): platforms and features are created (Section 3), measurements are uploaded (Section 4) and data is exported (Section 5). In Figure 1 the general setup of the Phenotype Database and the relation between the different elements is shown.

Figure 1: General overview of elements within the Phenotype Database. One subject (indicated by '1') can have more than one (number = N) sample(s) and one sample can have more than one (number = N) measurements in the same assay.

2 Generic Study Capture Framework section: set up your study

In this section it is explained in short how to set up your study, how to add subjects, define events and create samples in the Generic Study Capture Framework (GSCF) section of the Phenotype Database. For more details, please refer to the extended user manual.

2.1 Create a new study

1. Go to menu ‘Create’ and select ‘Create a new study’.
2. Select a predefined template or choose the add/modify option to create a new template.
3. Choose ‘General Study template’ in case you are not sure of the fields you need. This template contains the fields ‘title’, ‘description’, ‘code’, ‘startdate’ and ‘contacts’. Help text is available under the question mark icons.
4. Complete the mandatory fields indicated by a green arrow.
5. Click on ‘Quick save’.

2.2 Importing subjects using an Excel file

1. Go to the menu ‘Import’ and select ‘A part of the study design’.
2. Make sure that the Excel spreadsheet you use contains unique subject names.
3. ‘Page 1: Import file’. Click on “upload” to select the Excel file with subject information.
    a. Leave the date format: dd/MM/yyyy
    b. Select sheet containing data
    c. Indicate columns header (row number)
    d. Select data type being uploaded, here ‘Subject’
    e. Select study
    f. Select type of the ‘data template’. If no appropriate template is available, select ‘add/modify’ and ‘create new template’
    g. For a subject template two fields are mandatory: ‘subject identifier’ and ‘species’. Add other variables to describe subjects as desired (e.g. gender, age, tumour pathology, etc), but when using systems such as e.g. OpenClinica take care not to completely copy data already present within this tool.
    h. Click next
4. ‘Page 2: Assign properties’. Match fields in the subject template to the column headers present in the Excel file (‘name (IDENTIFIER)’ field should be matched to the Excel column with the subject identifier etc.). Click ‘next’.
5. ‘Page 3: Check the imported data’. Overview of the to-be-imported subjects. Check variables and where necessary make corrections/additions.
6. ‘Page 4: Confirmation’. If everything has been matched all right, click next, otherwise, click previous and make changes.
7. ‘Page 5: Done’. Subjects have been imported to the study.

2.3 Define (sampling-)events

In this part you need to state what happened to the subject, i.e. what events (e.g. treatment with a drug, diet challenge) or sampling-events (e.g. biological marker was measured on material from a subject, e.g. from a blood sample).

Go back to study for which you’ve imported subjects.

1. Go to the ‘Browse’ menu and choose ‘Edit’ your study.
2. Click on step 3 ‘Events’.
3. Choose the type of event (treatment or sampling event) and an ‘Event template’. Both for events and sampling-events, predefined templates are present.
4. If no appropriate template is present, select ‘add/modify’ and create a new (sampling-) event template.
5. Click add.
6. Continue adding all events you need.
7. Below a description will be given of a Group. Change the name of the Group to be more informative (e.g. name of measured biological marker).
8. Click on ‘quick save’ and then ‘next’.
9. An overview of subjects and sampling-events is now created. Mark the subjects that belong to the created (sampling-)event group.
10. Click ‘quick save’ and then ‘next’.

2.4 Samples

After defining your events (see section 2.3), an overview screen will appear with the to-be-generated sample-names (the name consists of the ‘subject ID’, ‘Event name’, ‘Group’ and ‘sampling-event template’). Depending on the sample template used additional fields will be present (which contains information on the sample type), which can be completed here.
Click ‘next’.

2.5 Assays

To be able to link your measurements to a sample for a particular sample-event, you need to create a so-called assay.

1. Select an assay template.
2. Give this assay a meaningful name (e.g. name of the measured biomarker).
3. Choose the appropriate module for this measurement (e.g. 'qPCR')
4. Click ‘add’.
5. Continue until you have defined all needed assays.
6. Click ‘quick save’ and then ‘next’.
7. An overview screen appears, showing all sample names for all subjects and the defined assays. Check the boxes to indicate which samples belong to what assay. Click ‘next’.

You have now created a study with subjects, sampling-events, samples and assays, which are awaiting data. To be able to upload the data you need to define platforms and features, which will be explained in section 3.

3 Creating platforms and features

Go to the ‘Module’ menu and select the module section which you want your assay to appear in (e.g. 'qPCR'). Now you need to choose (or create) a platform (section 3.1) and then features (section 3.2).

3.1 Creating a platform

Platforms have been predefined in the module section. A new platform may be created to distinguish between types of techniques, e.g. qRT-PCR SYBR Green.

To create a new platform:

1. Select Browse.
2. In case no appropriate platform is available, select ‘Create new platform’.
3. Complete the fields : ‘Name’, ‘Platform version’, ‘Platform type’, ‘Comments’ and ‘Template’. In case no appropriate predefined ‘Platform template’ is available, select add/modify and create the needed fields for the platform. Click ‘create’.
4. Click ‘back to list’.

Please note that it is possible to choose a different ‘platform type’ (referring to the different modules available) here. In case you have created a platform but can’t find it under the expected module, please check the other module and edit the ‘platform type’ for the platform.

3.2 Creating a feature

A feature describes the actually measured item (e.g. a biological marker, laboratory parameter etc) and the associated metadata for this measured item. It is important to add as much information to this feature as required in order to reproduce or trace back the results (according to ‘minimum information’ guidelines such as MIQE). Features should be present for all types of measurements in your study and should be uploaded to the appropriate platform. You can manually define features in Phenotype Database (section 3.2.1) or import them using an Excel file (section 3.2.2).

3.2.1 Manual creation of a feature

1. Go to the ‘Module’ menu, choose the appropriate module.
2. Select in the ‘Browse’ menu the option ‘Features’.
3. All current features are shown.
4. If no appropriate feature is available yet, select ‘create new feature’.
5. Indicate the platform to which this new feature belongs.
6. Give the new feature a meaningful name (mandatory) and if applicable a unit.
7. Select a feature template (if applicable for that template additional fields will now appear).
8. In case no appropriate template is available as yet, select add/modify and create a new feature template.
9. Each additional field should describe a specific characteristic for the measurement. Items ‘name’ and ‘unit’ are already mandatory, but other fields need to be added describing main features of

the experiment, whereby it is heavily recommended to report these features according to 'minimum information' guidelines.

3.2.2 Importing of features

When creating multiple features of the same type of measurement, importing of features will provide a quick way of uploading them to the module section. Of note should be that the descriptive field items themselves cannot be imported in this manner, but should be created beforehand . However, the *content* of these fields can be created in Excel and uploaded in one session for multiple field items. See Appendix I for an example.

1. Go to the 'module' menu, choose the appropriate module.
2. Select in the 'Import' menu the option 'Features'.
3. Select the Excel file containing information on the features.
4. Choose the appropriate data template (= Feature template). It is indicated that the data template is not required, but leaving this empty results in the use of the standard feature template. This will give problems in the next step as the descriptive fields from the Excel file cannot be matched.
5. Select platform. Click next.
6. Match the headers in the Excel file to the descriptive fields present in the selected feature template. Discard all columns and rows that do not need to be imported. Click next.
7. Check whether everything was matched correctly. If not, use the 'prev' to go back to the previous step. Click next.
8. You can now use these features to add measurements to your study.

4 Importing data/measurements

At the moment the Phenotype Database only handles the import of numeric values. In case you have data in text format, dichotomise that data if possible, by coding them in the Excel file (for instance Yes=1 No=0; unmethylated (U) = 0, methylated (M) = 1).

Go from the module back to the GSCF by clicking on 'Go to GSCF'.

1. Choose from the 'Browse' menu the 'My studies' (or 'All studies') option.
2. Click on the 'view' icon for the appropriate study.
3. Go to the tab 'Assays'. You will now see an overview of all assay names, modules, a column 'Link' (with a hyperlink to the eventual measurement details) and the sample names.
4. Click on details of the respective assay. No measurements are present as yet.
5. Click on the word 'importer' in the text 'Use the **importer** to import your data' to upload results into the assay.
6. Select the correct assay. Click next.
7. Select the correct platform.

8. Select the Excel (or CSV) file containing the results/measurements. Please make sure that the setup of the file to be imported is correct. Click on ‘allowed layouts’ text for examples. Click next.
9. Select sample layout (‘sample’ or ‘subject’). Click next.
10. Check that samples were matched to the correct subject, features and event, otherwise make adjustments here or click on previous if applicable.
11. Click next if all data have been matched correctly. Your data has been added to the assay and can be viewed and exported.

5 Export

You can export data from Phenotype Database in order to perform calculations or visualisations on the data, or to obtain an export file for upload into, for example, the data-integration platform tranSMART.

5.1 Export study (meta) data

1. Go to the GSCF.
2. Go to the ‘Export’ menu, choose ‘Export studies to SimpleTox file’.
3. Select the study of interest and click ‘Export’. The downloaded file will contain information on subjects, sampling-events and sample-names.

5.2 Export study (meta) data and measurements.

1. Go to the ‘Export’ menu and choose ‘Export assay data to file’.
2. Select your study and the assay.
3. In the next screen, mark the items of interest It is also possible to export only the measurements.
4. Choose the type of export file (in most cases ‘European (.csv)’) and click on ‘submit’.
5. Click ‘ok’ in the next screen to download. The downloaded file will contain information on subjects, sampling-events, sample-names, event groups and the measurements.

Appendix I: Example of fields within a qRT-PCR feature

For qRT-PCR measurements the following additional fields describing a real-time quantitative experiment should be added to a feature:

1. Target gene full name(s): full name(s) of the target gene, e.g. NCBI definition.
2. Target gene symbol (HGNC): name of the target gene using HGNC term (for human subjects)
3. Target gene ontology: NCBI code, GI code
4. Target gene forward primer: forward primer sequence; orientation given as standard: 5'-3' (in nucleotides)
5. Target gene reverse primer: reverse primer sequence; orientation given as standard: 5'-3' (in nucleotides)
6. Annealing temperature of target gene primers (in °Celsius)
7. Target gene primer manufacturer
8. Reference gene full name(s): full name(s) of the reference (housekeeping or control) gene, e.g. NCBI definition.
9. Reference gene symbol (HGNC)
10. Reference gene ontology: NCBI code, GI code
11. Reference gene forward primer: forward primer sequence; orientation given as standard: 5'-3' (in nucleotides)
12. Reference gene reverse primer: reverse primer sequence; orientation given as standard: 5'-3' (in nucleotides)
13. Annealing temperature of reference gene primers (in °Celsius)
14. Reference gene primer manufacturer
15. Reporter dye (+ manufacturer)
16. Description: description on how the qRT-PCR was performed. Can be reference to a published article or a reference to the protocol. It is also possible to enter a field with experimental protocol and upload the protocol file itself.
17. Machine used for qRT-PCR (+ manufacturer)
18. Software used for analysis of results (software, manufacturer and version)
19. Analysis/calculation + unit value of the measured feature